Interplay between harmonics and formants in singing : when vowels become music

M. Castellengo[a] and N. Henrich Bernardoni[b]
[a]LAM/d'Alembert, 11 rue de Lourmel, 75015 Paris, France
[b]GIPSA-lab, 11 rue des Mathématiques, 38402 Grenoble, France
michele.castellengo@upmc.fr

In human speech, the production of vowels consists in strengthening some specific areas of the harmonic spectrum, known as formants, by adjusting vocal-tract acoustical resonances with articulators such as tongue, lips, velum, jaw, and larynx. In singing, a compromise is often sought between the frequency of harmonics and resonance frequencies, sometimes at the expense of vowel perception. In some vocal cultures, this link between harmonic frequency and resonance frequency is skilfully adjusted. A melody is generated independently of the tonal melody related to vocal-fold vibrations. This is the case of harmonic singing, overtone singing or Xhoomij, practiced in Central Asia, but also of singing by Xhosa women in South Africa. In this paper, the adjustments between harmonics and formants are explored on a wide range of commercial singing recordings and experimental recordings in laboratory. Three main strategies are described from both acoustical and musical point of view. In a first case, the spectral melody is produced by a play on the first formant (F1). The first harmonic frequency is often kept constant and at low values due to period doubling induced by a ventricular vibration. In a second case, the spectral melody is produced by a play on the second formant (F2), with a higher frequency of the first harmonic. Complex spectral melody can also be developed by a vocal game on the first two formants. In particular, we will illustrate and discuss the cases where the two first formants evolve while remaining in an octave ratio (F2 = 2F1).

1 Introduction

When producing vowels in speech and singing, the fluid-structure interaction between air expelled from the lungs and moving walls induces vocal-folds vibration. This vibration generates a harmonic acoustic source, which propagates through the vocal tract (laryngeal and pharyngeal cavities, mouth and nasal cavities). The vocal-tract area function from glottis to lips is controlled by the speech articulators (tongue, lips, jaw, velum, larynx), which contributes to the adjustment of vocal-tract resonances (Ri). The resonances shape the harmonic voiced sound spectrum, in boosting acoustical energy in frequency bands designated in acoustics by the term *formants* (Fi). The frequency ratio between the first two formants F1 and F2 is perceptually coded into vowels.

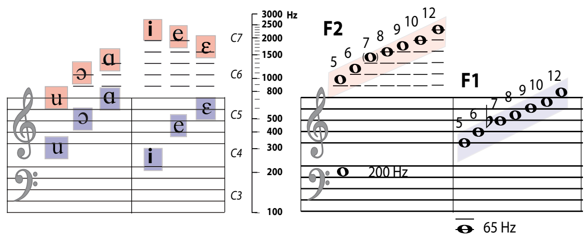

Figure 1: Mean values of formant frequencies F1 (blue) and F2 (red) on a musical scale. On left panel, the vowels have been grouped for which the two formants vary conjointly.

Several singing techniques illustrate harmonic-resonance adjustments. Possible interactions depending on sung pitch are shown in Figure 1, which presents the mean values of the two first formant frequencies for a male speaking voice. The vowel location on the diagram is only indicative. It depends on individual peculiarities and the chosen language. Besides, values are given for male speech, as the songs studied here are mainly produced by male singers. The first formant F1 ranges from 300 Hz (/i/) to 800 Hz (/a/), which corresponds on a musical scale to E4-G5. It covers the high range in male voices, the medium and high range in female voices.

In western classical singing, a tuning between the vocal-folds vibratory frequency (f0 = H1) and vocal-tract first-resonance frequency (R1) is sometimes mandatory to allow a loud and comfortable voice production, such as in the case of soprano high range [1, 2, 3] or, more generally when the sung pitch gets close to R1 [3]. To find a good balance between resonance adjustments and clarity of vowels constitutes a great part of the classical singer's training. Such singers have to be able to sing a text on a wide range of pitches. In traditional Croatian folk singing [4], in Bulgarian women's singing [5] or in Broadway Musicals [6], a systematic tuning is observed between the second harmonic (H2=2f0) and R1 for those vowels which do not have a too low first-resonance frequency. This practice gives power and clarity to the voice. It is produced by means of vowels /o/ /ɔ/ /ɛ/ /a/ in a limited pitch range: 220 to 320 Hz for male singers, 350-500 Hz for female singers (see Figure 2).

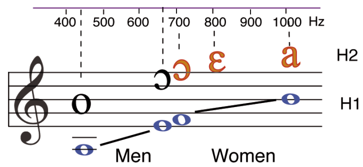

Figure 2: Illustration on a musical scale of vowels and pitches for which a tuning R1:2f0 is possible. The blue notes present the musical pitches.

The second formant F2 ranges from 600 Hz for vowel /u/ to 2400 Hz for vowel /i/ within the musical range E5-E7 (see Figure 1). Glottal fundamental frequency may come close to resonance frequency only for low-F2 vowels such as /u/ and /o/. In most cases, F2 lies well above f0, and it globally contributes to the voice quality. F2:f0 tunings have been observed in the soprano high range [2]. But most F2:Hi (i>1) tunings observed in the literature are reported for techniques of harmonic singing, which we shall now address. The literature will first be briefly reviewed. The tuning strategies will then be discussed on the basis of a wide range of commercial recordings. These observations will be supplemented by a case study of a Mongolian singer by means of simultaneous acoustical recordings and ultrasound observations of tongue motion.

2 Harmonic singing : the state of the art

A spectral melody and low-pitch tone - In the singing techniques mentioned above, a melody is produced by varying the vocal-folds vibratory frequency and the resonances are tuned depending on vowel and sound quality. Roles are reversed in harmonic singing. A great surprise may be felt at first while listening to a diphonic singing. In 1847, the singer and singing teacher Manuel Garcia (son) [7] reported that he heard the Baskirs having the astonishing ability to produce altogether two distinct parts: a pedal and a high-pitched melody (« *l'étonnante faculté de produire à la fois deux parties distinctes : une pédale et une mélodie aiguë »)*. In a cultural tradition where a wind instrument produces a single pitch, it seemed unexplainable that a single man could simultaneously produce two clearly identifiable sounds. Phonatory adjustments developed in singing skilfully contribute to this effect. First laryngeal sound is produced on a constant pitch, so that listener's attention is soon deviated from it to the harmonic spectral variations. Similar principle is found in mouth harp or musical bow, yet perceptual effect is differently explained. The vibrating blade of mouth harp or the string of musical bow generate a sound of stable frequency, which is perceived as dissociated from the musician. Or any voice production, even on a sustained tone, has a frequency jitter on the order of 3%. The two acoustic phenomena in harmonic singing, a low-pitch drone sound and a spectral melody, are correlated in frequency by the jitter. Consequently, they are attributed to the same singer, which explains the listener's stupefaction and the use of the term "diphonic singing" to describe it

History of scientific studies - The amazement caused in western cultures by the « bitonal » effect of these songs has aroused several scientific studies. The first acoustical data were gathered on Tibetan chant [8]. The precision of H5 and H10 harmonic selection was evidenced on a vowel /o/ sung at very low pitch (75 Hz, F1). It was also observed that the singer started his song at 150 Hz (F2) before switching to the octave below, indicating a way of producing voice sound still unknown.

With improvement of acoustical analyses and laryngeal visual examination techniques, the research studies have gone in two directions. Some studies focused on the harmonic selectivity which characterizes these songs [9, 10, 11]. Others focused on the period-doubling phenomenon [12, 13, 14, 15]. These academic studies are based on singing production of non-native singers from classical tradition who were introduced to diphonic singing. Thus they may not fully reflect the reality and the diversity of traditional vocal practices, that only ethnomusicologists truly investigate [16, 17, 18, 19, 20]

The most recent research studies deal with vocal-tract modelling, in taking into account the peculiarities of these songs: sygyt [21, 22], khöömei [23, 24], kargiraa [25]. However, several open questions remain unanswered about musical aspects of this vocal art which covers a large tessitura, and even more the perceptual questions it raises [26]. In this paper, we aim to study the main realizations which are as many original solutions to answer the question : how "to play a melody" with vowels?

3 Material and methods

Database - Songs extracted from a wide range of commercial recordings have been used to explore the relationship between tonal melody and formantic one. Table 1 lists the song samples. Additional information is given about the country of origin, the mean fundamental frequency (f0), the range in F1 and F2, and the technique used (period-doubling, usual phonation in M1 or M2).

Table 1: Songs database (see Discography section for references on commercial recordings and singers). *Mong* : Mongolia. *Mech* refers to the laryngeal mechanism used by the singer (M1, M2) or the use of ventricular-folds vibration (f/2 is period doubling).

Code	origin	melody	mech	f0(Hz)	F1	F2
K1	Tibet	F1	f/2	68	340	---
K2	Tibet	F1-F2	f/2	75	375 H5-10	Octave
K3	Tuva	F1-F2	f/2	62	370-750 H6-H12	Octave
K4	Tuva	F1-F2	f/2	61,75	400-750 H8-H12	Octave
K5	Tuva	F1-F2	f/2	65,5	518-880 H8-H12	Octave + ornaments
K6	Mong	F1-F2	f/2	68,5	400-826 H7-H10	Octave & ornaments
K7	Mong	F1-F2	f/2	75	518-880 H7-H12	Octave
K8	Mong	F1-F2	f/2	68	550-820 H8-H12	Octave
K9	Mong	F1	f/2	76	320-920 H2-H5	680-1700 H8-H12
K10	Tuva		f/2	92	570-840 H6-H9	1040-2050 H11-H22
K11	Tuva	F1	f/2	73,5	444-888 H6-H12	1600-1900 (…)
X1	Mong	F2	M1	207	?	1250-2500 H6-H14
X2	Mong	F2	M1	214	?	1700-2600 H8-H12
X3	Mong	F2	M1	200	?	1200-2600 H8-H12
X4	Mong	F2	M1	208	?	1600-2500 H8-H12
X5	Tuva	F2	M1	212	?	1700-2800 H8-H13
X6	Tuva	F2	M1	190	?	1520-2280 H8-H12
X7	Mong	F2	M1	165	?	1100-1800 H8-H11
F	Xhosa	F1-F2	f/2	110 122	350-650 H3-H6	Octave

The singer K9 was recorded in laboratory, by means of synchronized audio and tongue-motion ultrasound-based imaging [27]. The singer was sitting down in an anechoic chamber. The ultrasound transducer was maintained in contact to the skin below the chin by means of a head and transducer support. The tongue motion was visualized in a sagittal plane. The tongue tip and root may be hidden by respectively jaw and hyoid-bone acoustic shadow. The data were acquired and processed in Matlab (Ultraspeech, [27]).

Musical characteristics of harmonic singing - Traditional singers aim at producing a recognizable melody, such as the one that they play with their vielle or sing with their ordinary voice. In studying numerous harmonic singing, we notice that most musicians use the harmonic range from the 6th harmonic to the 12th one. These harmonics correspond to the successive degrees of a pentatonic scale and to the main musical intervals (quart, major and minor thirds, tones). The singer can use either the first resonance or the second one to boost the selected harmonics in the adequate frequency range. Right panel in Figure 1 shows comfortable pitches for tuning the resonances to the selected harmonics: G3 for F2:H6-H12 and C2 for F1:H6-H12. However, from a perceptual point of view, both resonances may boost the harmonic spectrum and be heard. We can thus question the role played by the other resonance than the one used for harmonic melody, and in particular the musical interval which separates them.

4 F2-based spectral melody

Musical and acoustical description - The F2-based songs are the most perceptually astonishing. They are also the most simple ones to analyze. The study of 7 songs (X1-X7) shows that singer ‘s fundamental frequency is found between 160 Hz and 220 Hz, the median frequency being 200 Hz (G3). The example analyzed in Figure 3 is particularly clear and controlled. The spectral melody is performed between H5 and H12, so between 1000 and 2400 Hz.

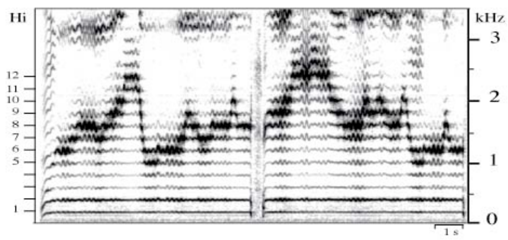

Figure 3: F2-based song in Sygyt style (X3)

These F2-based songs are characterized by a lack of energy on voice fundamental frequency, a missing first formant, and a precise melodic line thanks to a great harmonic selectivity. Voice is produced in laryngeal mechanism M1. The sound is perceived as very loud, as the harmonic melody evolves within a frequency range close to frequency band of maximal auditory acuteness. All these factors contribute to the perception of a melody standing out of the vocal sound and being whistled and superimposed to the low-pitch drone sound. The F1 neutralization accounts for the fact that no vowel is truly perceptible.

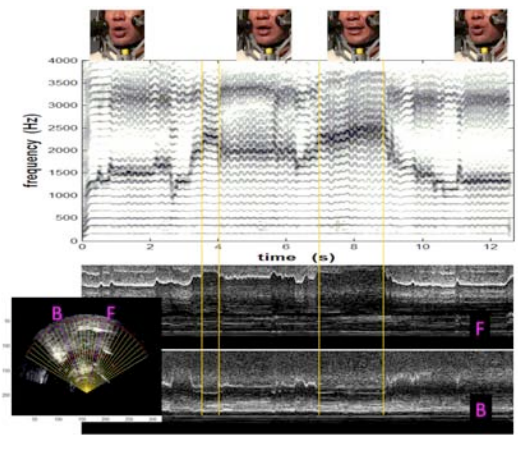

Figure 4: Analysis of a F2 song sample (K9). The top panels present lip aperture along the sequence, the middle panel the sonagraphic analysis. Bottom panels present the tongue kymograms for two polar lines selected at the front (F) or back (B) parts of the tongue body.

Tongue-motion observations - The first radiographic investigations (X-rays) [28, 19] and resonance magnetic imaging [24] conducted on three singers show the main role of the tongue tip, which is raised and moved very close to the hard palate, behind the superior teeth. Small adjustments of the mouth and lips go together with the observed tongue motion. Two singers move the tongue forward to produce the most high-pitched sounds [29, 19] while one moves it backward [24].

As illustrated in Figure 4, singer K9 adjusts exactly his lip aperture for each sound, simultaneously to his tongue height. In the sagittal plane visible in ultrasound, the front part of the tongue remains flat and its height position is controlled very accurately.

5 F1-based spectral melody

Musical and acoustical description - As shown in Table 1, all the songs were F1 is varied are produced with a low-pitch drone sound (period doubling). The analysis of 11 F1-based songs (K1-K11) led to a separation in two groups. For the lowest ones, fundamental frequency is found around 65 Hz (C2), with the use of H6 (450 Hz) to H12 (900 Hz). Fundamental frequency in the second group is on average 82 Hz (Eb2). The spectral melody does not go beyond H10.

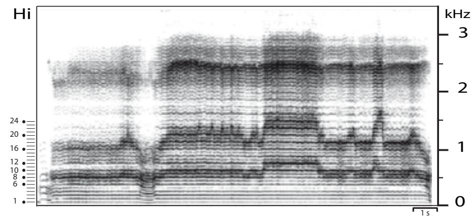

Figure 5: F1-F2 based song in kargyraa style (K3)

Figure 5 illustrates a song in kargiraa style (f0=62 Hz), for which period doubling comes clearly visible at the attack. The harmonic spectrum is rich, and several formantic bands are visible. The song frequently starts on a common pitch (130Hz), and the switch to period doubling follows as soon as the aerodynamical coupling between vocal-folds and ventricular-folds vibration is stabilized [8,15]. Sound is perceived as very loud, hoarse. The perception of sonorous vowels is prominent at first listening, which stands in the way of perceiving a spectral melody. The ease to perceive it depends on the singer’s ability to tune F1 with great accuracy.

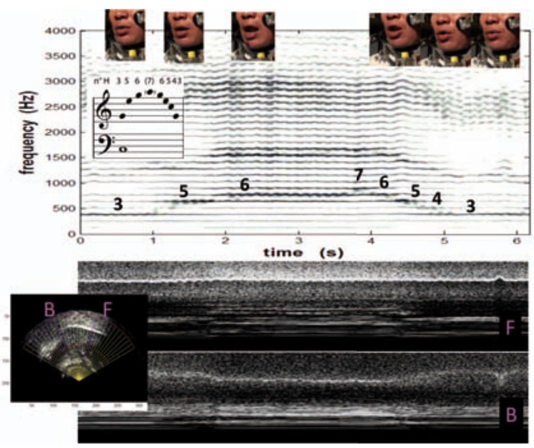

Figure 6: Analysis of a song sample where F1 and F2 vary in octave (K9). The top panels present lip aperture along the sequence, the middle panel the sonagraphic analysis and musical notation. Bottom panels present the tongue kymograms for two polar lines selected at the front (F) or back (B) parts of the tongue body.

Tongue-motion observations - A flattened tongue is observed in the literature [24, 19]. For one singer, the tongue moves to the hard palate together with the raise in harmonic melody. Mouth is more widely opened than for F2-based songs. In Figure 6, we present an analysis in which F1 and F2 evolves at the same time, close to the octave. Polar kymograms of ultrasound-based tongue motion reveals an immobility of the tongue front part. The video observation of the mouth indicates that the harmonic-formantic tuning is controlled by jaw and lips movements.

6 F1-F2 spectral melody enhanced

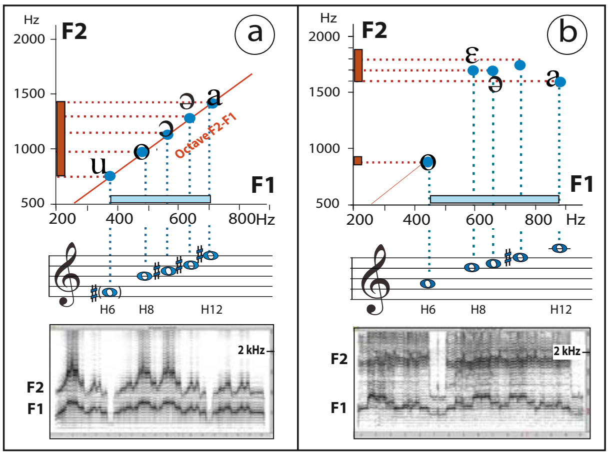

Figure 7 : a) « Formantic octave » song (K4) ; b) F1-based song with F2 kept constant (K11)

The traditional F1-based songs (Kargyraa) being always rich in harmonics, F1 and F2 can simultaneously be perceived and the vowels are very salient. In some vocal traditions – such as the songs "*A Tenore* " of Sardinia – vowels are used to enrich the harmonic content of a spectral melody. But a perceptible melody is produced if either the spectral accuracy of F1 is enhanced or the amplitude of F2 is reduced, which will darken the voice timbre. Two strategies will now be discussed, which are noteworthy from a perceptual point of view.

Octave ratio between formants : F2 = 2F1

Some singers develop a technique which consists in selecting simultaneously and with great accuracy the harmonics related to F1 and F2 which are on an octave ratio. They only make use of open vowels /o/, /ɔ/, /ɑ/, /a/. Bottom left panel in Figure 7a shows the analysis of a Tuva song, and top left panel the F1 / F2 vocalic plane. The line F2 = 2F1 is plotted on the vocalic plane, and the corresponding vowels are given. This tuning technique confers increased brightness to the spectral melody, without cancelling the perception of vowels.

Formant F2 kept constant

Figure 7b shows another option to favour the perception of F1 spectral melody. Singers succeed in stabilizing F2 frequency by means of vowels /ɛ/, /ə/, /a/. F2 is kept constant and far enough from F1 within the frequency band 1600-1800 Hz, which additionally confers intensity and brightness to the kargyraa sound

7 The special case of Xhosa women's singing in South Africa

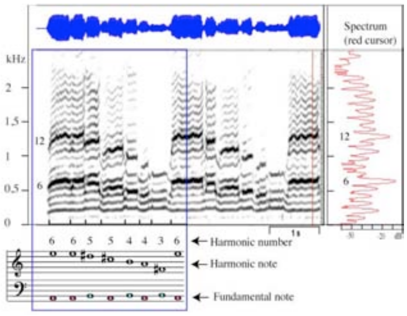

Figure 8 : « Formantic octave » song with two alternation fundamental frequencies.

Few women practice harmonic singing. Yet Xhosa women in South Africa develop such technique, combining several original styles of harmonic singing with period doubling [17]. The song illustrated in Figure 8 is produced in the style *umngqokolo ngomqangi*, as an imitation of musical bow. To compensate for the fact that their fundamental frequency in laryngeal mechanism M1 does not reach very low frequencies, the Xhosa singers have developed an original singing technique in producing alternatively two fundamental frequencies separated by a tone. This technique allows them to access a complete musical scale. By means of harmonics H4, H5, H6 at f0=110 Hz and harmonics H3, H4, H5 at f0=122 Hz, they can sing within a system similar to the one that they would have between harmonics H6 to H12 if their fundamental frequency was f0=55 Hz. Similarly to the kargiraa singer in Figure 7a, they sing a spectral melody where the two first formants are in octave ratio by making use of vowels between /ɔ/ and /u/.

8 Conclusion

The notion of harmonic singing covers a wide range of musical performances. All these performances are original solutions to a three-constraints problem: the use of vocalic formants to produce a harmonic melody independently of voice fundamental frequency; the tuning of voice fundamental frequency to one or the other of the two formantic bands so as to benefit from H6 to H12 for the desired musical scale; the perceptual deviation of vowels for melodic purposes. Some vowels are skillfully chosen, in particular the open vowels which can even be used to sing with formants F1 and F2 in octave ratio. Many aspects remain to be explored in more details, in particular the timbre richness. The few spectrographic analyses combined with tongue and lip motion recorded on a Mongolian singer call for further analyses of the coordination between these movements and the acoustical and musical achievement.

Acknowledgments

We are very grateful to our research colleague, Thomas Hueber, thanks to whom the ultrasound recordings were made. We thank Bayarbaatar Davaasuren, the renowned Mongolian singer who participated to the experiment, and Bernard Fort, who introduced us to him.

Discography

CD "Inédit Mongolie" - Auvidis, W 260009 (1989), tracks: 4 (X1); 5 (X2; X7); 6 (X3).

CD "Voices from the center of Asia" - Smithsonian Folkways, SF 400017 (1990), tracks: 1 (K5); 4 (X5); 9 (K11); 14 (K10; X6); 18 (K4).

CD "Les voix du monde", CNRS-Harmonia mundi, CMX 374 1010.12 (1996), CD-II-37 (K3).

CD "The Heart of Dharma", Ellipsis Arts (1996), track 2 (K3).

Dave Dargie demonstration tape, track A-1 (F).

Alash Ensemble - Singers : Bady Dorzhu-Ondar (K6; K7; K8); Kongar-ool Ondar (X4).

Bayarbaatar Davaasuren, (2013), Gipsa-Lab (K9).

Data from H. Smith (1967), lama from the Gyutu Monastery near Dalhousie, recorded in 1964 (K2).

References

[1] E. Joliveau, J. Smith and J. Wolfe, “Vocal tract resonances in singing: The soprano voice”, *J. Acoust. Soc. Am.* 116 (4), 2434-2439 (2004)

[2] M. Garnier, N. Henrich, J. Smith, J. Wolfe, « Vocal tract adjustments in the high soprano range, *J. Acoust. Soc. Am.* 127 (6), 3771-3780 (2010)

[3] N. Henrich, J. Smith, and J. Wolfe, “Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones”, *J. Acoust. Soc. Am.* 129 (2), 1024-1035 (2011)

[4] P. Boersma and G. Kovavic, “Spectral characteristics of three syles of Croatian folk singing”, *J. Acoust. Soc. Am.* 119 (3), 1805-1816 (2006)

[5] N. Henrich, M. Kiek, J. Smith, and J. Wolfe, “Resonance strategies in Bulgarian women's singing”, *Logopedics Phoniatrics Vocology* 32, 171-177 (2007)

[6] T. Bourne, M. Garnier, “Physiological and acoustic characteristics of the female music theater voice”, *J. Acoust. Soc. Am.* 131 (2), 1586-1594 (2012)

[7] M. Garcia jr, “Mémoire sur la voix humaine; réimpression augmentée de quelques observations nouvelles sur les sons simultanés”, p.24, Paris: Duverger (1840)

[8] H. Smith, K.N. Stevens and R.S. Tomlinson, “On an unusual mode of chanting by certain Tibetan lamas”, *J. Acoust. Soc. Am.* 41 (5), 1262-1264 (1967)

[9] G. Bloothooft, E. Bringmann, M. Van Cappellen, J.B. Van Luippen, et al. “Acoustics and perception of overtone singing” *J. Acoust. Soc. Am.* 92 (4), 1827-1836 (1992)

[10] F. Klingholz, “Overtone singing: productive mechanisms and acoustic data”, J. of Voice 7 (2), 118-122 (1993)

[11] H. K. Schutte, D.G. Miller and J.G. Švec, “Measurement of formant frequencies and bandwith in singing”, *J. of Voice* 9 (3), 290-296 (1995)

[12] L. Dmitriev, B. Chernov and V. Maslow, “Functioning of the Voice Mechanism in Double Voice Touvinian Singing”, *Folia Phoniatrica* 36, 193-197 (1983)

[13] L. Fuks, B. Hammmarberg and J. Sundberg, “A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences”, *TMH-QPSR3,* 49-59 (1998)

[14] J. G. Švec, H. K. Schutte and D. G. Miller, “A subharmonic vibratory pattern in normal vocal folds”, *J. of Speech and Hearing Research* 39, 135-143 (1996)

[15] L. Bailly, N. Henrich and X. Perlorson, “Vocal fold and ventricular vocal fold vibration in period-doubling phonation: physiological description and aerodynamic modeling”, *J. Acoust. Soc. Am.* 127 (5), 3212-3222 (2010)

[16] A.N. Askenov, “Tuvin folk music”, *Asian Music* 4 (2), 7-18 (1973)

[17] D. Dargie, “Xhosa music: its techniques and instruments, with a collection of songs”, Cape Town: David Philip

[18] H. Zemp and T. Q. Hai, “Recherches expérimentales sur le chant diphonique”, *Cahiers d’ethnomusicologie* 4, 27-68 (1991)

[19] T. C. Levin and M. E. Edgerton, “The Throat Singers of Tuva”, *Scientific American* 218 (3), 70-77(1999) and related video files (X-rays)

[20] J. Curtet, “La transmission du *höömij,* un art du timbre vocal : ethnomusicology et histoire du chant diphonique mongol”, Thèse de doctorat, Université de Rennes 2.

[21] M. Kob, “Analysis and modeling of overtone singing in the sygyt style”, *Applied acoustics* 65 (12), 1249-1259 (2004)

[22] C. Tsai, Y. Shau and T. Hsiao, “False vocal fold surface waves during Sygyt singing: A hypothesis”, *Proc. ICVBP,* (2004)

[23] S. Adachi and M. Yamada, “An acoustical study of sound production in biphonic singing, Xöömij”, *J. Acoust. Soc. Am.* 105 (5), 2920-2932 (1999)

[24] K.-I. Sakakibara, H. Imagawa, T. Konishi, K. Kondo et al, “Vocal fold and false vocal fold vibrations in throat singing and synthesis of Khöömei”, *Proc. ICMC,* (2001)

[25] P. Lindestad, M. Södersten, B. Merker and S. Granqvist, “Voice source characteristcs in Mongolian “throat singing” studied with high-speed imaging technique, acoustic spectra, and inverse filtering”, *J. of voice* 15 (1), 78-85 (2001)

[26] P. Cosi and G. Tisato, “On the magic of overtone singing”, *Voce, Parlato. Studi in onore di Franco Ferrero,* 83-100 (2003)

[27] T. Hueber, G. Chollet, B. Denby, M. Stone, “Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application”, Proc. of ISSP, 365-369 (2008)

[28] H. Zemp and T.Q. Hai, “Le chant des harmoniques”, film 16 mm, Paris: Musée de l’Homme and CNRS-AV [http://videotheque.cnrs.fr/doc=606], (1989)